4-2-2020

Bootstrap Statistics

Greg Hutto
University of Arkansas, Fayetteville

Engineering and Operations Management Lunch & Learn Webinar Series

PRESENTS

"Bootstrap Statistics"

April 2, 2020

Online Degree Options

Expand breadth and depth of engineering knowledge	*Provide leadership and business skills to manage technology teams*	*Improve effectiveness and efficiency of operations*
M.S. in Engineering	**M.S. in Engineering Management**	**M.S. in Operations Management**
MSE Comprehensive Exam	MSEM Comprehensive Exam	MSOM Comprehensive Exam
3 Electives chosen from MSEM, OMGT, and Engineering courses	3 Electives chosen from MSEM, OMGT, and Engineering courses	6 Electives chosen from OMGT and MSEM courses
4 core courses from approved list Computer Applications Mathematics Management Technical Communications	4 core courses EMGT 5033 Intro to Engineering Mgt OMGT 5783 Project Management OMGT 5463 Economic Decision Making INEG/OMGT 5443 Decision Models	4 core courses OMGT 5003 Intro to Operations Mgt OMGT 5123 Finance or 5463 Economic Decision Making OMGT 5623 Strategic Management OMGT 5783 Project Management
Approved 3 course engineering sequence to form cohesive topic area	Approved 3 course engineering sequence	Undergraduate Prereqs (if required) OMGT 4853 Decision Support Tools OMGT 4323 Industrial Cost Analysis OMGT 4333 Statistics OMGT 4313 Law and Ethics
ABET Accredited Bachelors Degree in Engineering		Any Regionally Accredited Bachelors Degree

TODAY'S PRESENTER

Greg Hutto

Test Wing Chief Operations Analyst

*Mr. Gregory Hutto serves the 96th Test Wing as Chief Operations Analyst, responsible for embedding designed experiments as the principal test method for several hundred tests each year. He teaches an extensive series of short courses in test methods to all testers in the Wing from the Wing Commander to our 600+ scientists and engineers. As a LtCol in the USAF Reserves, he served as senior military advisor to AF Operational Test & Evaluation Center and as special advisor for test design to the Flight Test Center commander at Edwards AFB, California. Mr. Hutto is a distinguished graduate of the US Naval Academy in Operations Research and holds a Master's in the same field from Stanford University. Over the past 30 years, he has served in nearly every branch of test and evaluation from laboratory basic science to joint operational field testing. He would like to publicly repent of his 11 years of testing without the benefits of the principles of well-designed experiments.

Statistics by Experiment – The Spirit of Resampling

- Take out a coin – we're gonna procreate!
- Flip twice for a family of two children
  - Heads=Male; Tail=Female
  - Pls vote your families in poll: MM MF FM FF
- Pr(exactly one boy and and one girl)
  - Monte Carlo Experiment – what do our flips say?
- Pr(other is a boy|one child is a boy)
  - Monte Carlo Experiment – what do our flips say?

	A	B	C	D	E	F	G	H
1		Door						
2	GameNum	1	2	3	Prize Door	You Choose	Switch	Stay
3	1	A New Car!	*goat*	*goat*	1	2	Win	*Lose*

The Monte Hall Problem

Behind one of three doors is a "NEW Car!!"; behind the other two are goats. You must pick a door. After you pick, Monte Hall opens one door and <always> shows you a goat. Before he shows you the car, Monte asks if you want to stay with your original door or switch to the other unopened door.

Do you
STAY
SWITCH or
it DOESN'T MATTER ???

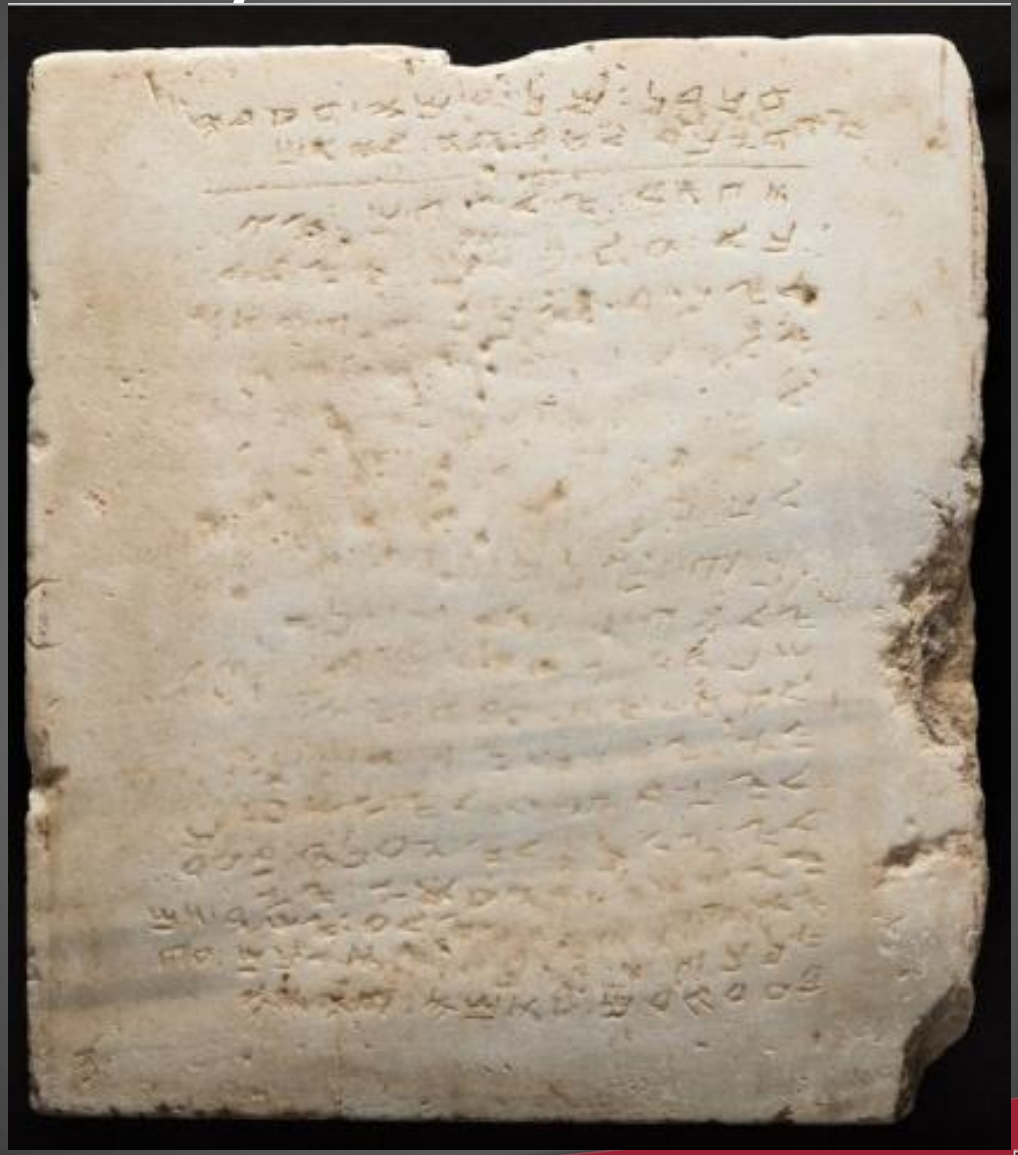
- Monte Hall experiment setup -- we'll play 1000 games
- What do you think? Please vote:
  Switch-Stay-Doesn't Matter
- Answer later at the end of the seminar

Basic Stats Course Circa 1976 -- P&S 101

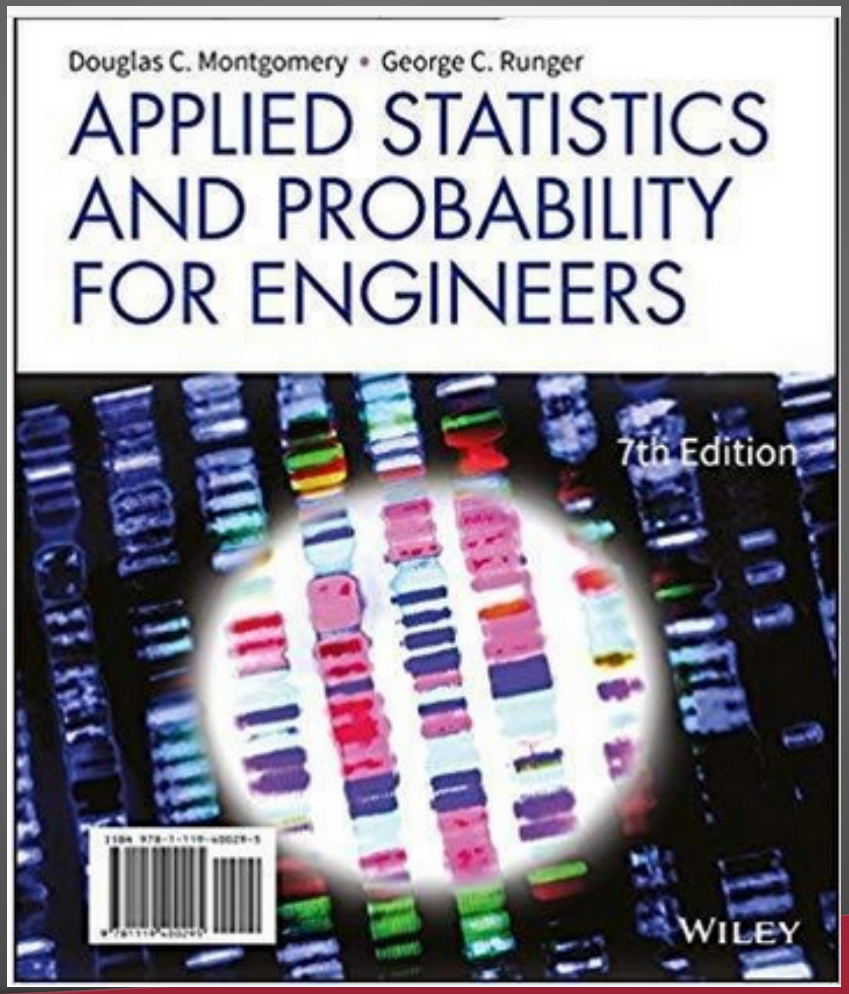
Chapters followed this outline:

7. Sampling Distributions and Point Estimates for Parameters
8. Statistical Intervals for Single Samples
9. Single Sample Hypothesis tests (inferences)
10. Statistical inference for two samples
    - ❑ Differences in means
    - ❑ Differences in variances
    - ❑ Paired t Test
    - ❑ Differences in proportions

My class notes…



Most Prob & Stats are taught this way … but is there an alternative?



Estimating a number and its precision

Elves Frolic … but how much??

ESTIMATION. Deep in the Forest of Nod the Elves frolic alone. In a recent Frolic Fest 20 elves imbibed the elixir Elsinore and danced and frolicked an average of 30 hours before needing another Elsinore to recharge. The elves are righteously proud of their frolicsome fortitude. But not all frolic equally! How strong is the evidence for 30-hour elves? Might the average be substantially more or less in future fests? Tim the Enchanter is called in to adjudicate…

There are some who call me …Tim!

The data and questions

- We have a single estimate of the average frolic time
- What might future averages look like? How big? How small?

	Frolic-er	Elves
	1	33
	2	28
	3	30
	4	34
	5	24
	6	28
	7	29
	8	27
	9	32
	10	31
	11	28
	12	30
	13	26
	14	28
	15	34
	16	36
	17	31
	18	29
	19	34
	20	28
Parameters	Averages	30.0
	St Dev	3.1
	Median	29.5
	Range	12
	90th Decile	34
	Skewness	0.19

We need more information!

- An obvious answer is to gather more samples …
- BUT, if I assume the original sample represents the population – shape, center, dispersion …
- How about I just re-use <u>this</u> sample?

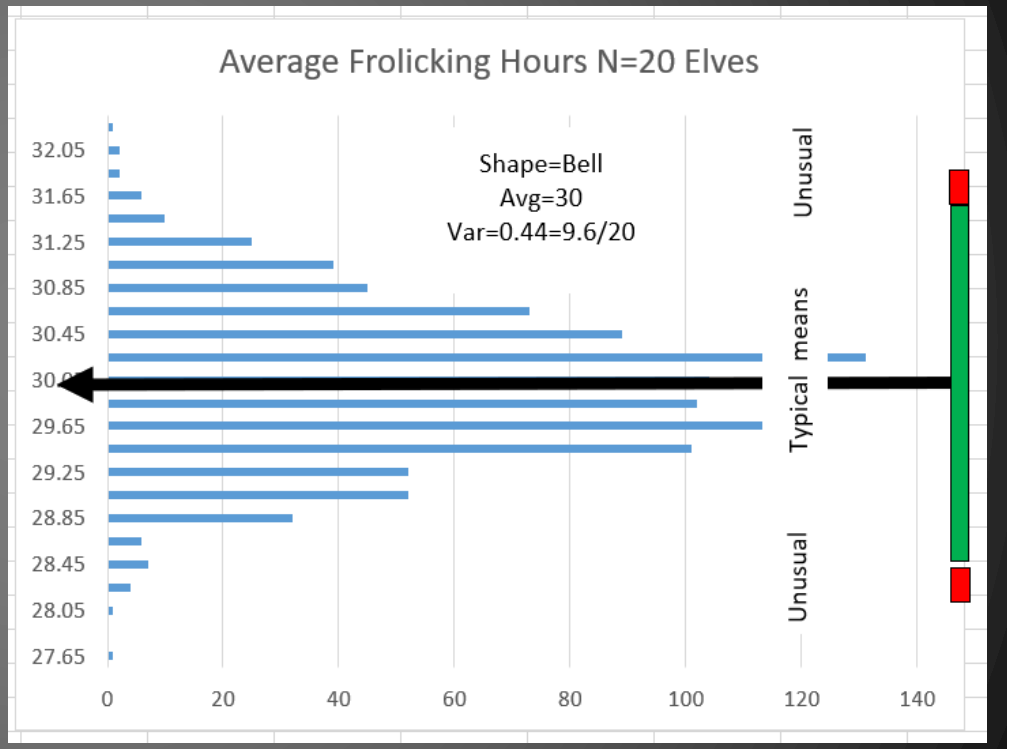
	Frolic-er	Elves
	1	33
	2	28
	3	30
	4	34
	5	24
	6	28
	7	29
	8	27
	9	32
	10	31
	11	28
	12	30
	13	26
	14	28
	15	34
	16	36
	17	31
	18	29
	19	34
	20	28
Parameters	Averages	30.0
	St Dev	3.1
	Median	29.5
	Range	12
	90th Decile	34
	Skewness	0.19

Random re-samples from original sample w/replacement								
Frolic-er	1	2	3	4	5	6	000	1000
1	34	28	29	34	26	30		24
2	28	34	33	26	28	28		36
3	32	28	28	34	34	26		28
4	31	31	29	33	27	28		31
5	34	30	33	34	28	29		28
6	33	36	29	34	29	31		30
7	27	33	28	28	36	29		28
8	28	32	36	29	33	31		29
9	28	28	28	28	34	31		28
10	28	29	34	36	28	29		31
11	28	29	28	29	26	31		31
12	28	28	29	36	24	31		26
13	27	31	24	34	28	28		28
14	33	29	28	31	26	36		34
15	28	28	34	31	28	26		34
16	28	28	30	28	31	31		27
17	32	36	28	36	31	32		27
18	33	32	27	32	29	34		28
19	36	30	36	32	29	28		31
20	29	36	29	26	34	28		24
Averages	30.3	30.8	30.0	31.6	29.5	29.9		29.2

Draw 1000 re-samples of 20 w/ replacement & track the average

Result – histo (distribution) of 1000 samples (with replacement)

- This mean distribution can stand-in for results of more sampling
- Sampling distribution defines typical vs. unusual frolic averages
- Central Limit Theorem, anyone?

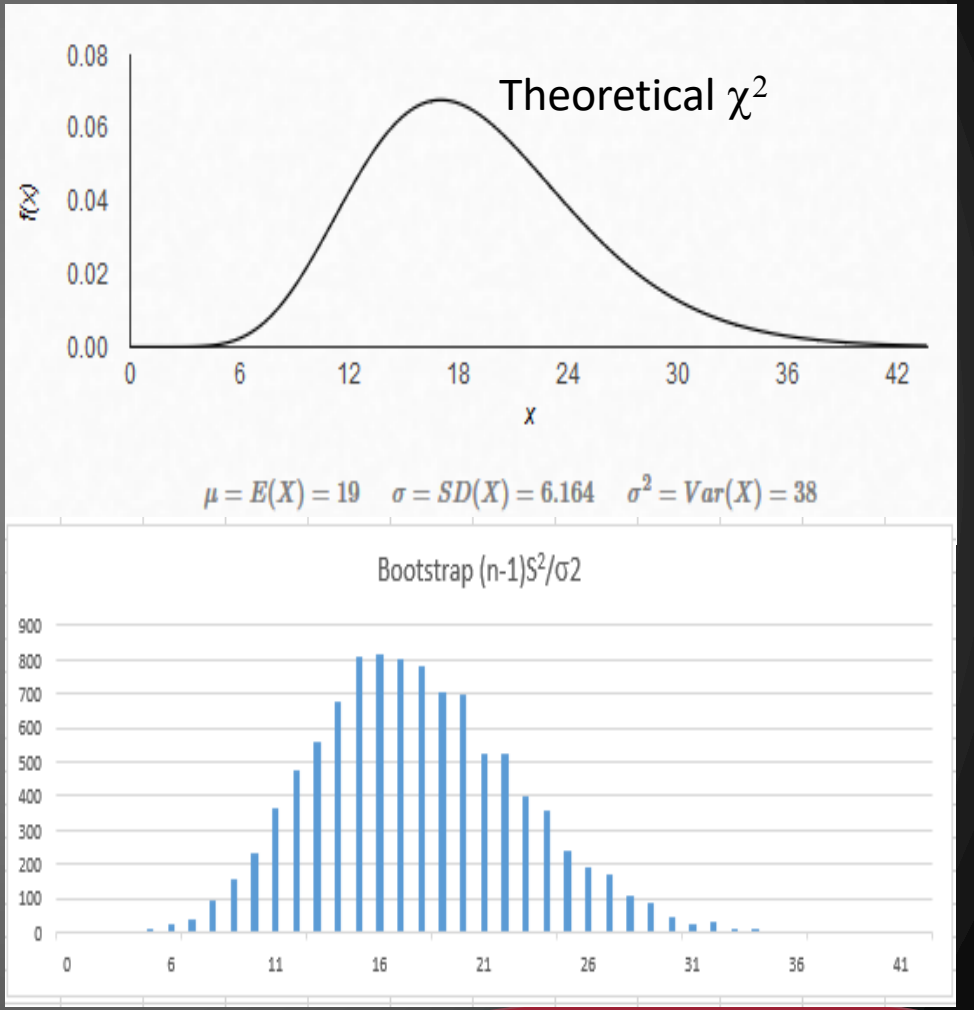


We have a mean point estimate & a related sampling distribution

We can estimate *any* sampling distribution

- We might recall the sampling distribution of the Variance is a function of a chi square (χ^2(n-1)) … (or not)
- We can bootstrap the sampling distribution of…
  - Proportions
  - Medians
  - Ranges
  - Other percentiles

  Any function!

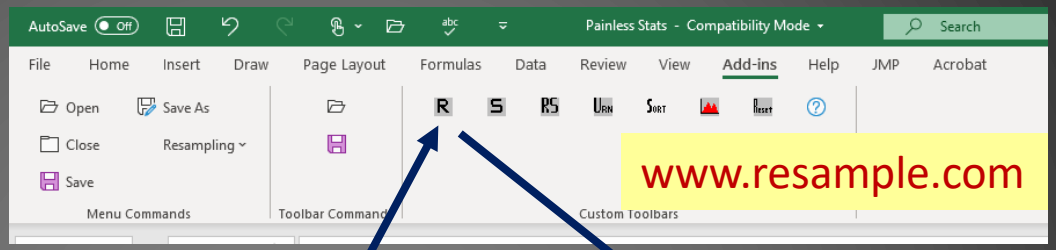


UNIVERSITY OF ARKANSAS

Quick Resampling Mechanics – rsxl Resample.com

Algorithm

1. Question to answer
2. Function that answers it
3. How to resample (w/ or w/o replace)
4. How many resamples



	Frolic-er	Elves
	1	33
	2	2
	3	30
	4	34
	5	24
	6	28
	7	29
	8	27
	9	
	10	
	11	8
	12	30
	13	26
	14	28
	15	34
	16	36
	17	31
	18	29
	19	34
	20	28
Parameters	Averages	30.0
	St Dev	3.1
	Median	29.5
	Range	12
	90th Decile	34
	Skewness	0.19

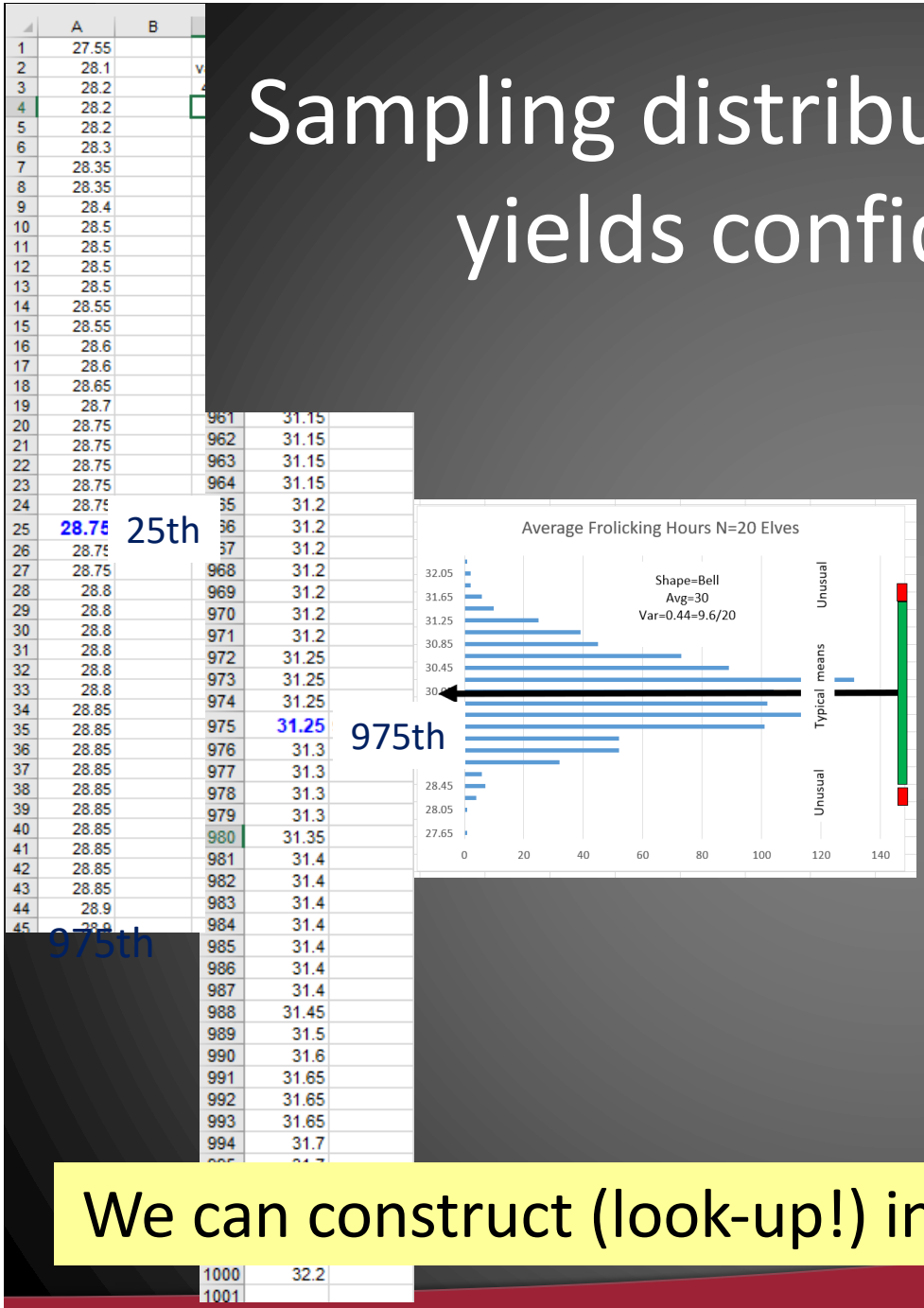
Frolic-er	Elves
1	28
2	30
3	31
4	28
5	31
6	34
7	28
	28
	30
	26
	28
12	28
13	28
14	31
15	31
16	24
17	34
18	31
19	28
20	34
Averages	29.6

Template for Re-samples

"Score Cell" to track

UNIVERSITY OF ARKANSAS

Sampling distribution of 1000 means yields confidence intervals



- Cut off 2.5% on the bottom & top (25 values each)
- It would be surprising if the true mean of 20 elves was outside [28, 32]
- 95% of the time (28.75, 31.25) contains the mean
- BTW can also look up prediction & tolerance intervals

We can construct (look-up!) interval estimates for the mean

Historical Resampling Timeline[1]

1650	Gambling Experiments – Blasé Pascal
1700	Probability Theory – the four Bernoulli brothers
1850	
1900	Student t Distribution W.S. Gosset
1925	The Tea Experiment – R.A. Fisher Fisher's Exact Test (permutations) – R.A. Fisher
1950	Monte Carlo approaches (RAND et al.) Maurice Quenouille (and later Tukey's) Jackknife
1975	Applications to Business and Economics – Julian Simon
1980	Publication of Efron's (Stanford) Article on Bootstrap
1990	Peter Hall's Publication of *Asymptotic Theory of Resampling*
2000	

1. Source: Chapter 1 *Bootstrap Methods and Their Application*, AC Davidson, 1997

Is Bootstrap a Fraud?

THE TRAVELS
AND
SURPRISING ADVENTURES
OF
BARON MUNCHAUSEN
ILLUSTRATED WITH
THIRTY-SEVEN
CURIOUS ENGRAVINGS
FROM
THE BARON'S OWN DESIGNS
AND FIVE ILLUSTRATIONS
By G. CRUIKSHANK

Source: www.munchasuen.com

"Much to my chagrin, I found myself at the bottom of the lake!" declared the Baron.

"But Baron, how did you save yourself from a watery fate?" exclaimed his lady listener.

"Why, I simply reached down," the Baron explained " and pulled myself up by my bootstraps."

Gosset Resampled to Generate the Student t Distribution

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

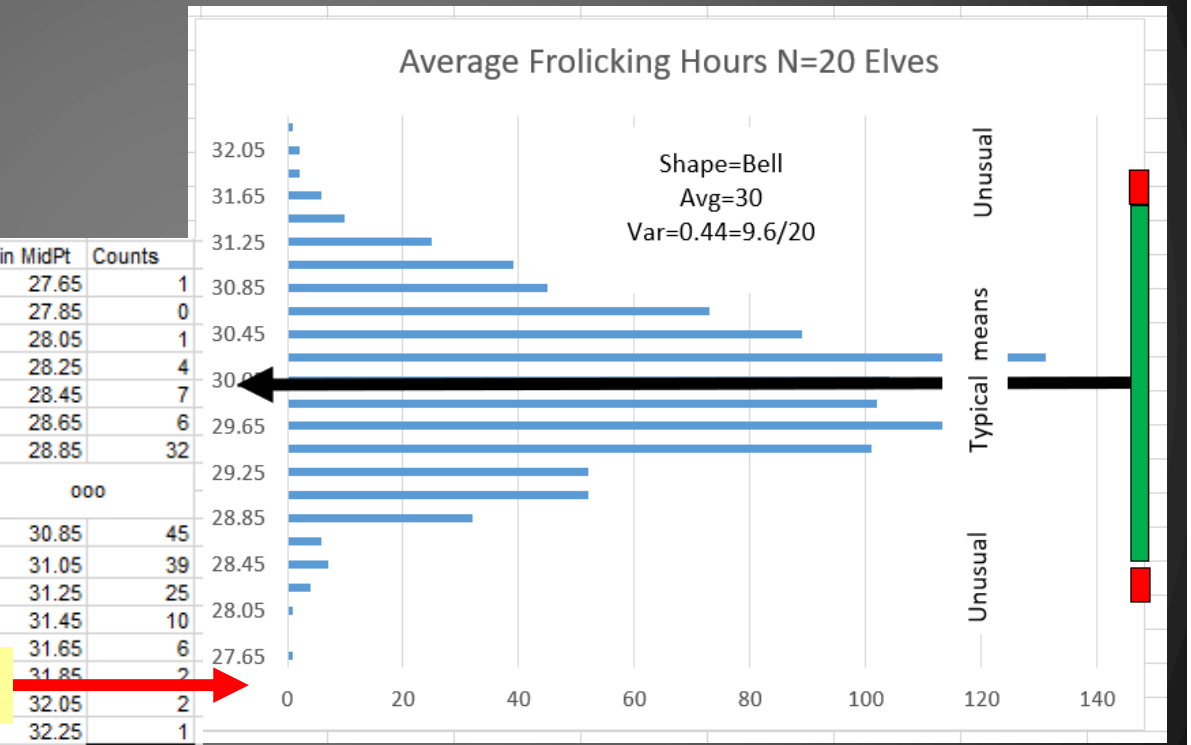
Testing a Claim about a Number

New question – suppose the elves had imbibed NEW Elsinore™ Elixer an al! & improved formulation … can we support a claim that Elsinore extends frolic-time to 32 hours?

Like the standard normal (z) or student t, we can look a score up in this distro

- Claim (u0 = 32)
- Look-up how often see claim or more extreme?
- 3 in 1000
- Is this p?
- Single sample t* = -2.90 / p=0.005)
- Thumbs-down on New Elsinore™

Bin MidPt	Counts
27.65	1
27.85	0
28.05	1
28.25	4
28.45	7
28.65	6
28.85	32
ooo	
30.85	45
31.05	39
31.25	25
31.45	10
31.65	6
[illegible]	2
32.05	2
32.25	1
	1000



Don’t need no formula, tables, Excel functions, Statpack

Testing a Claim About Two Samples

Elves and Sprites Frolic …

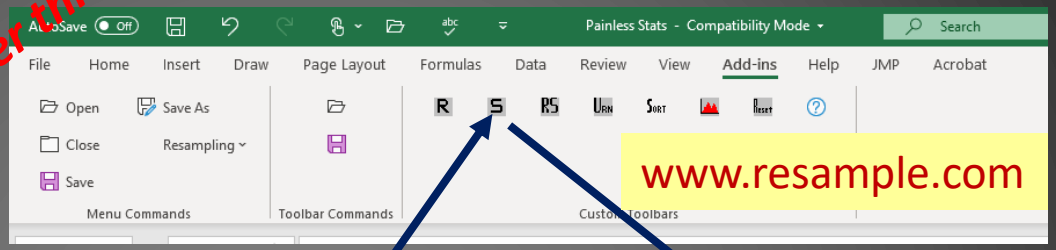
but who is *more* frolicsome?

Two Sample Resampling Mechanics

Algorithm

1. Question do Elves frolic 2.4 hours more? "real"??
2. Function that answers it
3. How to resample (w/ or w/o replace)
4. How many resamples

Remember this



Frolic-er	Elves	Sprites
1	33	29
2	32	28
3	31	30
4	28	
5	31	
6	33	26
7	29	22
8	24	27
9	30	27
10	28	31
Averages	29.9	27.5
Difference	2.4	

Frolic-er	Elves	Sprites
1	24	29
2	24	29
3		26
4		28
5		31
6	27	31
7	31	28
8	31	33
9	27	22
10	32	28
Averages	28.9	28.5
Elf-Fairy	0.4	

Template for Shuffles (w/o replacement)

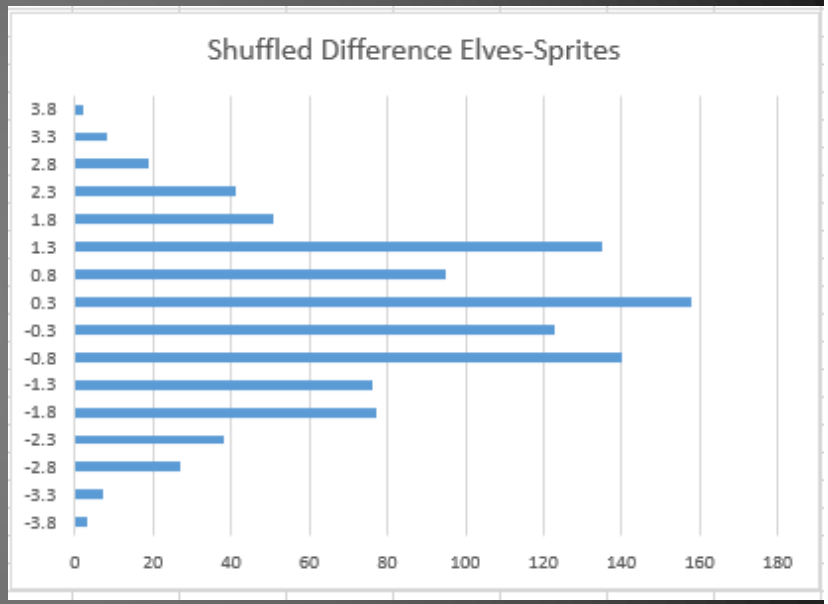
"Score Cell" to track

Like the standard normal (z) or student t, we can look a score up in this distro

- Claim (means equal)
- Look-up how often see observed result or more extreme?
- Count 44 in 1000
- Is this p?
- Two sample t* = -1.88 / p=0.0.039)
- Elves Rule!

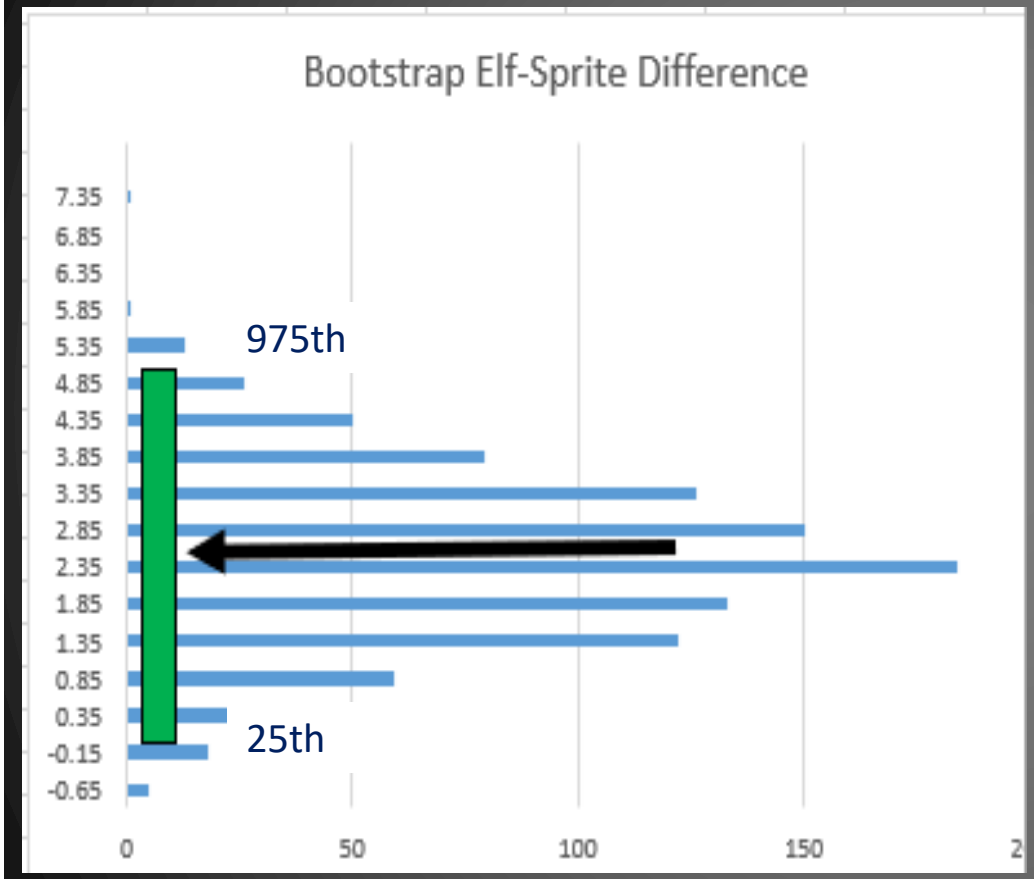
2.4

Bin MidPt	Counts
3.8	2
3.3	8
2.8	19
2.3	41
1.8	51
1.3	135
0.8	95
0.3	158
-0.3	123
-0.8	140
-1.3	76
-1.8	77
-2.3	38
-2.8	27
-3.3	7
-3.8	3



Don’t need no formula, tables, Excel functions, Statpack

Sampling distributions yield confidence intervals: column means, difference



- NOW sample *within* columns (why?)
- Cut off 2.5% on the bottom & top (25 values)
- 95% of the time (0.2, 4.8) contains the mean difference – 0 not included
- Plus prediction & tolerance intervals

We can construct (look-up!) interval estimates

We've covered our Basic Stats Course

Chapters followed this outline:

- ✓ Sampling Distributions and Point Estimates for Parameters
- ✓ Statistical Intervals for Single Samples
- ✓ Single Sample Hypothesis tests (inferences)
- ✓ Statistical inference for two samples
  - ✓ Differences in means
  - ✓ Differences in variances
  - ✓ Paired t Test
  - ✓ Differences in proportions
- ✓ Plus Intervals

Definitions

Bootstrap: Statistics done under the assumption that the population distribution is represented by the sample distribution. The data are sampled with replacement, simulating an infinite population from the sample.

Permutation (Shuffling) Methods: The basis for Fisher's Exact test and the Tea Experiment. The gold standard of resampling schemes – sampling is without replacement.

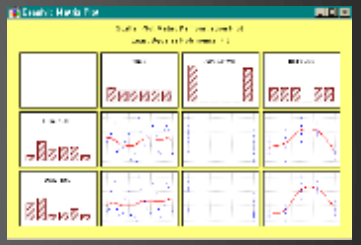
Resampling: Statistics done under computation-intensive methods involving sampling from some distribution. These include bootstrap, jackknife, data shuffling schemes, and possibly others.

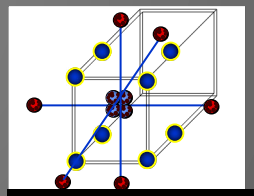
Monte Carlo: Simulations done using random (or pseudo-random) numbers. These might include resampling schemes, but for our purposes today, we'll distinguish between them.

Teaching Difficult Statistical Concepts via Resampling

- "Statistically Significant Differences" - Inferences
- Estimating Percentiles
- α and β errors
- Confidence intervals
- Sample size via OC curves
- Correlation coefficients
- Regression coefficients

			High Altitude		Low Altitude	
ECM	Offset	Clutter	Auto Trk	Man Trk	Auto Trk	Man Trk
Wet	Short	High	12.1			5.6
		Low		14.7	4.6	
	Long	High		12.2	7.8	
		Low	13.6			3.4
Dry	Short	High		12.1	5.6	
		Low	14.7			4.6
	Long	High	12.2			7.8
		Low		13.6	3.4	

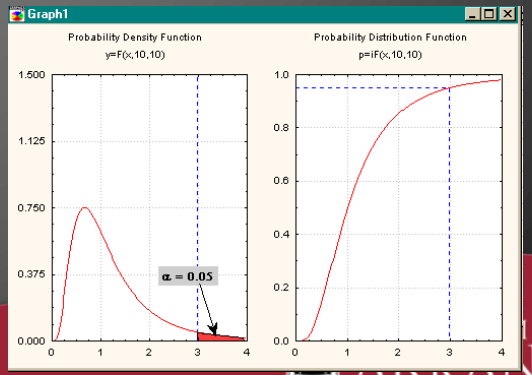




ANOVA; Var.:LOCERRPC; R-sqr=.37782; Adj:.2285

EXPERIM. DESIGN: 2**(3-0) design; MS Residual=10450.88
DV: LOCERRPC

Factor	SS	df	MS	F	p
(1)THREAT	39466.7	1	39466.67	3.776398	.063321
(2)BOUND	7602.1	1	7602.08	.727411	.401821
(3)JINK	32864.8	1	32864.76	3.144689	.088359
1 by 2	61.6	1	61.56	.005890	.939436
1 by 3	78618.9	1	78618.86	7.522705	.011100
2 by 3	45.0	1	45.02	.004307	.948194
Error	261271.9	25	10450.88		
Total SS	419930.8	31			



Resampling crystallizes thinking

- Why must α be set to compute β error?
- What is the root problem behind missing data and unbalanced designs?
- What assumptions are we making about the sample when we apply the t test?
- What are the real effects of outliers on linear models?

Great Tool but No Panacea -- Drawbacks & limitations

- Inadequate sample size
- Inadequate resamples
- Nonrepresentative data
- Bias in difficult parameters to estimate

Bootstrap References & Resources

- See papers and links uploaded – email me for more
- Bradley Efron http://statweb.stanford.edu/~ckirby/brad/other/
- Julian Simon http://www.juliansimon.com/
- John Tukey http://people.bu.edu/aimcinto/jackknife.pdf
- Resample.com -- Excel Add-in I'll demo in a moment
- R – bootstrap and shuffle codes
- Can do within Excel with Index & Random functions

http://www.real-statistics.com/non-parametric-tests/resampling-procedures/

Summary of Bootstrap – Statistics Without Pain

- The Bootstrap (Resampling) provides a method that is:
  - Simple to teach and use
  - Resistant to misuse
  - Robust against violations of assumptions
  - Helps with missing data
  - Dealing directly with the problem
  - Illustrative of difficult statistical concepts
- Add the Bootstrap to your toolbag today

Julian Simon – Author
Resampling Stats

Monte Hall Problem -- The Answer

	A	B	C	D	E	F	G	H
1		Door						
2	GameNum	1	2	3	Prize Door	You Choose	Switch	Stay
3	1	A New Car!	*goat*	*goat*	1	2	Win	*Lose*

The Monte Hall Problem

Behind one of three doors is a "NEW Car!!"; behind the other two are goats. You must pick a door. After you pick, Monte Hall opens one door and <always> shows you a goat. Before he shows you the car, Monte asks if you want to stay with your original door or switch to the other unopened door.

Do you
STAY
SWITCH or
it DOESN'T MATTER ???

- Results of 1000 games … **SWITCH!**

Countif!!	Stay	333	=COUNTIF(Stay,"Win")
	Switch	667	=COUNTIF(Switch,"Win")

UNIVERSITY OF ARKANSAS

Questions?

Type in Chat

Greg Hutto
ghutto@uark.edu

UNIVERSITY OF ARKANSAS

M.S. in Operations Management At a Glance:

- Online and Live Course Options
- 30 Credit Hours (10 Graduate Courses)
  - With up to 4 pre-requisite classes
- Five 8-week Sessions Per Year
- No GRE/GMAT required with 3.0 Bachelor's GPA
- Total Program Cost is $12,000 to $15,000 (depending on pre-reqs needed)
- Can be completed in one year, but you have up to six years to complete

UNIVERSITY OF ARKANSAS

Project Management Graduate Certificate At a Glance:

- Online and Live Course Options
- 12 Credit Hours (4 Graduate Courses)
- 8-week sessions
- Five Enrollment Periods: Aug, Oct, Jan, Mar, May
- Entire Program Cost: Approximately $5,000
- 2.5 GPA with a Bachelor's Degree required for admission
- Certificate courses can also count toward MSOM degree

Coming up April 16, 2020

Lunch-n-Learn Webinar Series

When Optimal Isn't Optimal

Come join us for a presentation on how to help your organization make the best decisions in response to disruptive tech-driven change.

Presented by Robin Burk

Speaker, bestselling author, consultant and an expert on the disruptive tech that is rapidly changing our world.

THANKS FOR ATTENDING!

- For information about our flexible degree program options, email Karin Hickenbotham at kahicken@uark.edu.
- The video from today's webinar will be available on our website, *registered* participants will receive an email with the video link.
- Registration is open for next webinar on our website